B.M.S. COLLEGE OF ENGINEERING BENGALURU
Autonomous Institute, Affiliated to VTU

An Internship Report

Document Chat-Bot

Submitted in partial fulfillment for the award of degree of

Bachelor of Engineering
in
Computer Science and Engineering

Submitted by:
Nithin S (1BM21CS120)

Internship Carried Out

At

SIEMENS

SIEMENS Technology Services and Private Ltd. Gold Hill Excelsior Park
RMXG+JJM, Phase 2, Electronic City,
Bengaluru, Karnataka 560100

Internal Guide

Seema Patil
Assistant Professor
BMSCE

External Guide

Kolanda Bharath Kumar
Team Lead
Siemens Technology

Department of Computer Science and Engineering
BMS College of Engineering
Bull Temple Road, Basavanagudi, Bangalore 560 019

2023-2024

B.M.S. COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We, **Nithin S (1BM21CS120**) students of 6th Semester, B.E, Department of Computer Science and Engineering, BMS College of Engineering, Bangalore, hereby declare that, this Internship entitled " Dataset Collection and Analysis ", has been carried out under the guidance of Kolanda Bharath Kumar, Team Lead, Siemens Technology India, Seema Patil, Assistant Professor, Department of CSE, BMS College of Engineering, Bangalore during the academic semester March - June 2024. I also declare that to the best of my knowledge and belief, the Internship report is not a part of any other report by any other student.

Signature of the Candidates

Nithin S (1BM21CS120)

B.M.S. COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Internship titled "**Document Chat-Bot**" has been carried out by **Nithin S (1BM21CS120)** during the academic year 2023-2024.

Signature of the Guide

Signature of the Head of the Department

Signature of Examiners with date

1. Internal Examiner _____
2. External Examiner _____

Abstract

The project "Document Chat-Bot" aims to develop an intelligent chat-bot capable of assisting users by retrieving, summarizing, and providing information from a variety of documents. This chat-bot leverages advanced natural language processing (NLP) techniques and machine learning algorithms to understand user queries and deliver accurate and contextually relevant responses. The primary objective is to enhance the efficiency and user experience in accessing and managing large volumes of documents. The chat-bot can be deployed in various domains, including customer service, technical support, and internal documentation management. This project was undertaken as part of an internship at Siemens under the department of Digital Asset and Service Solutions (DASS), providing practical experience in applying AI technologies to real-world problems.

INDEX

Chapter 1: About the Company

Siemens AG is a global powerhouse in industrial engineering and technology, founded in 1847 by Werner von Siemens. Headquartered in Munich, Germany, Siemens has grown into one of the world's leading technology companies, focusing on electrification, automation, and digitalization. The company's diverse portfolio includes products and solutions for power generation and transmission, automation and control, medical imaging and laboratory diagnostics, and infrastructure and building technology.

Major Milestones:

- **1847**: Siemens & Halske was founded.
- **1866**: Invention of the dynamo machine.
- **1969**: Development of the first automated high-speed train.
- **2011**: Siemens introduces the world's most efficient gas turbine.
- **2020**: Siemens completes the spin-off of Siemens Energy.

Siemens operates in more than 200 countries and employs over 300,000 people worldwide. In fiscal year 2022, Siemens reported a revenue of €87 billion. The company is organized into several business divisions, including Digital Industries, Smart Infrastructure, Mobility, Siemens Healthineers, and Siemens Energy. Each division focuses on specific market segments and technologies, driving innovation and growth in their respective areas.

Chapter 2: About the Company's Department

The Digital Asset and Service Solutions (DASS) department at Siemens is dedicated to leveraging digital technologies to enhance asset performance, optimize service delivery, and drive operational efficiency. DASS integrates advanced technologies such as the Internet of Things (IoT), artificial intelligence (AI), big data analytics, and cloud computing to develop innovative solutions that meet the evolving needs of customers across various industries.

Functions and Responsibilities:

- **Research and Development**: Continuous innovation in digital solutions for asset management and service optimization.
- **Solution Development**: Designing and implementing software solutions that enhance asset performance and service efficiency.
- **Data Analytics**: Utilizing big data and analytics to provide insights and predictive maintenance solutions.
- **Customer Support**: Offering technical support and training to ensure the effective implementation and use of digital solutions.
- **Collaboration**: Working closely with other departments and external partners to integrate solutions and drive digital transformation.

Organizational Structure:

The DASS department is structured into several teams, each specializing in different aspects of digital solutions. These teams include software development, data science, digital strategy, customer support, and project management. Each team is led by a team leader who reports to the department head. The department fosters a collaborative environment where cross-functional teams work together to deliver innovative solutions and achieve organizational goals.

Overall, the DASS department plays a critical role in Siemens' digital transformation journey, helping clients optimize their operations, reduce costs, and enhance overall efficiency through the use of cutting-edge digital technologies.

Chapter 3: Tasks Performed

During the internship, the primary task was the development of the Document Chat-Bot. The project involved the following key activities:

- **Requirement Analysis**: Collaborated with the project team to understand the functional and non-functional requirements of the chat-bot.
- **Design and Architecture**: Designed the system architecture for the chat-bot, including the NLP components, data storage, and user interface.
- **Development**: Implemented the chat-bot using Python and integrated NLP libraries such as Lang chain and Vector stores like FAISS.
- **Testing and Validation**: Conducted rigorous testing to ensure the chat-bot's accuracy and reliability in responding to user queries.
- **Documentation**: Prepared comprehensive documentation detailing the design, implementation, and usage of the chat-bot.

These tasks required extensive use of both technical and non-technical skills, including software development, problem-solving, teamwork, and effective communication.

Chapter 4: Reflection Notes

The internship at Siemens provided valuable hands-on experience in developing advanced AI-based solutions. Key technical outcomes include proficiency in NLP techniques, software development skills, and an understanding of system architecture design. Non-technical outcomes include enhanced communication skills, time management, and the ability to work effectively in a team environment. The experience has significantly contributed to personal and professional growth, providing a strong foundation for future career endeavors.

REFERENCES

- "Artificial Intelligence: A Modern Approach" by Stuart Russell and Peter Norvig.
- "Speech and Language Processing" by Daniel Jurafsky and James H. Martin.
- "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron.
- "Natural Language Processing with Python" by Steven Bird, Ewan Klein, and Edward Loper.
- "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Jacob Devlin et al.
- "Attention Is All You Need" by Ashish Vaswani et al.
- "Transformers in NLP: A Survey" by Qiu et al.

- Technical manuals and user guides provided by Siemens.
- Project-specific documentation and reports available within the Siemens DASS department.